Ch. 8 Math Preliminaries for Lossy Coding

8.4 Info Theory Revisited

Info Theory Goals for Lossy Coding

Again – just as for the lossless case – Info Theory provides:

- Basis for Algorithms & Bounds on Performance

The Info Theory Goal for Lossy: Given a probabilistic model for a class of signals, determine the best possible Rate-Distortion curve

Want the lower bound

- Want $D(R)$ vs. R ("Info Theory R-D" rather than "Operational R-D")

Recall from our Figure of Compression Processing:

- original signal $x[n]$ takes values on continuum
- recovered signal $y[n]$ takes discrete values

Raises Questions:

1. How much information is in $x[n]$?
2. How much of the info in $x[n]$ is conveyed by $y[n]$

Another way of asking #2 is: If I know $y[n]$, how much uncertainty remains about the signal $x[n]$?

Recall: Info - Uncertainty

Discrete-Discrete (D-D) Results

Discrete-Discrete - Motivation

Q#1 is made hard because $x[n]$ takes values on a continuum – as we'll see later. Let's sidestep this by first answering Q#2 for the case when

$x[n]$ and $y[n]$ ***both*** take discrete values

… but values from different sets

Example: Let the original signal values be $x[n] \in \{0, 1, 2, \ldots, 15\}$

Let the "compressed" signal be $y[n] \in \{0, 2, 4, \ldots, 14\}$

where the mapping is

$$\begin{array}{lll} y[n]=0 & if & x[n]=0 \; or \; 1 \\ y[n]=2 & if & x[n]=2 \; or \; 3 \\ y[n]=4 & if & x[n]=4 \; or \; 5 \\ Etc. & & \end{array}$$

If $x[n]$ is IID & equally likely then $H(X) = 4$ bits

… and it is easy to verify that $H(Y) = 3$ bits

This concept allows one to answer the following:

- If I know $y[n]$, how much uncertainty (i.e., unknown info) remains about $x[n]$?

- For our example: Once I know $y[n]$, I know that $x[n]$ is one of only two (equally likely) values…

- Thus, by "inspection": $H(X|Y) = -½\log_2(½) - ½\log_2(½) = 1$ bit

- Note: $H(X) - H(Y) = 4 - 3 = 1$ bit, also!

<u>Conditional Entropy (D-D): *H*(*X*|*Y*)</u>

Define Cond. Info of input <u>symbol</u> x_i… given output <u>symbol</u> y_j as:

$$i(x_i \mid y_j) = -\log_2 P(x_i \mid y_j)$$

Now, to assess the source as a whole – ***<u>on average</u>*** – we average over all possible x_i & y_j:

$$H(X \mid Y) = -\sum_{i=0}^{N-1} \sum_{j=0}^{M-1} P(x_i, y_j) \log_2 \left[P(x_i \mid y_j) \right]$$

Need joint Prob. to avg. over all possible pairs

Captures Cond. Info

Using $P(x_i,y_j) = P(x_i|y_j)\,P(y_j)$ gives

$$H(X \mid Y) = -\sum_{i=0}^{N-1} \sum_{j=0}^{M-1} P(x_i \mid y_j) P(y_j) \log_2 \left[P(x_i \mid y_j) \right]$$

Conditional Entropy of Source Alphabet *X* given Reconstruction Alphabet *Y* (and the rule that maps *X* to *Y*)

$H(X|Y)$ = Avg. Uncertainty on X given Y

= On Avg. how much info about X is ***NOT*** conveyed by Y

Clearly… $\boldsymbol{H(X|Y) \le H(X)}$

➔ Knowing Y can "help" but can't "hurt"!

Ex. 8.4.2 in the textbook computes $H(X|Y)$ for our example above and shows that $H(X|Y) = 1$… consistent with what we showed.

Avg Mutual Information of *X* & *Y* (D-D)

Now we answer Q#2: How much info about X is conveyed by Y (on average)?

Define "Avg. Mutual Info": $I(X;Y)$

It is obvious that: $$I(X;Y) = H(X) - H(X|Y)$$

- $I(X;Y)$: Amount of Info Y conveys about X
- $H(X)$: Total Info that could be conveyed
- $H(X|Y)$: Amount of Info that isn't conveyed

The uncertainty left unresolved about X

$I(X;Y)$ is the theoretical minimum rate we need to use to specify the information that Y carries about X

Can show that: $$I(X;Y) = I(Y;X)$$

Continuous-Discrete (C-D) Results

Intro to Continuous-Discrete Results

All of the above was pretty simple – but it was for the case where both $y[n]$ ***and*** $x[n]$ has ***discrete*** values!

When $x[n]$ takes values on a continuum none of this works! WHY?

Consider the interval [0,1] with the #'s occurring according to the uniform PDF..

If we want to represent every # in this interval using binary, how many bits do we need for each #?

Similarly, $i(x[n]) = -\log_2 P(x[n])$ Probability, not PDF!

The prob. of cont. RV taking any specific value is zero!

$i(x[n]) = \infty$ if $x[n]$ is on a continuum!

If we can't get $i(x[n])$… we can't get $H(X)$

… and we can't get $H(X|Y)$

… and we can't get $I(X;Y)$

Development of Info of Cont. RV

What do we do??!!! "Limits" to the rescue!!!!

Divide up the continuum into cells of width Δ…

(Yes… we are going to let Δ→0 !)

Cells: $[(i-1)\Delta, i\Delta)$

$$\exists\, x_i \in [(i-1)\Delta, i\Delta) \quad s.t.$$

For each cell, $$\underbrace{f_X(x_i)\Delta}_{\text{This \textbf{is} a probability}} = \int_{(i-1)\Delta}^{i\Delta} f_X(x)\,dx$$

PDF

Define <u>discrete</u> RV X_d that takes on the values x_i with prob. function

$$P(X_d = x_i) = f_X(x_i)\Delta$$

The entropy of X_d is then $$H(X_d) = -\sum_{i=0}^{N-1} P(x_i)\log_2\left[P(x_i)\right]$$

Now (in some sense) $\lim_{\Delta \to 0} H(X_d)$

should give the entropy of the cont. RV *X*

It is easy to show (see pp. 205-206 if you must!):

$$\lim_{\Delta \to 0} H(X_d) = \underbrace{\lim_{\Delta \to 0} \left[-\sum_{i=-\infty}^{\infty} [f_X(x_i) \log_2 f_X(x_i)] \Delta \right]} - \underbrace{\left[\lim_{\Delta \to 0} \log_2 \Delta \right]}_{= -\infty}$$

$$-\int_{-\infty}^{\infty} f_X(x) \log_2 f_X(x) dx$$

$$\triangleq h(X)$$

Called "**Differential Entropy**" of *X*

Causes Entropy of *X* to be infinite

Applying this reasoning to any cont. RV always gives that pesky lim log$_2$Δ!!!

***But*... each RV will have an *h*(*X*) that characterizes it!!!**

Ex.: Differential Info of Gaussian X

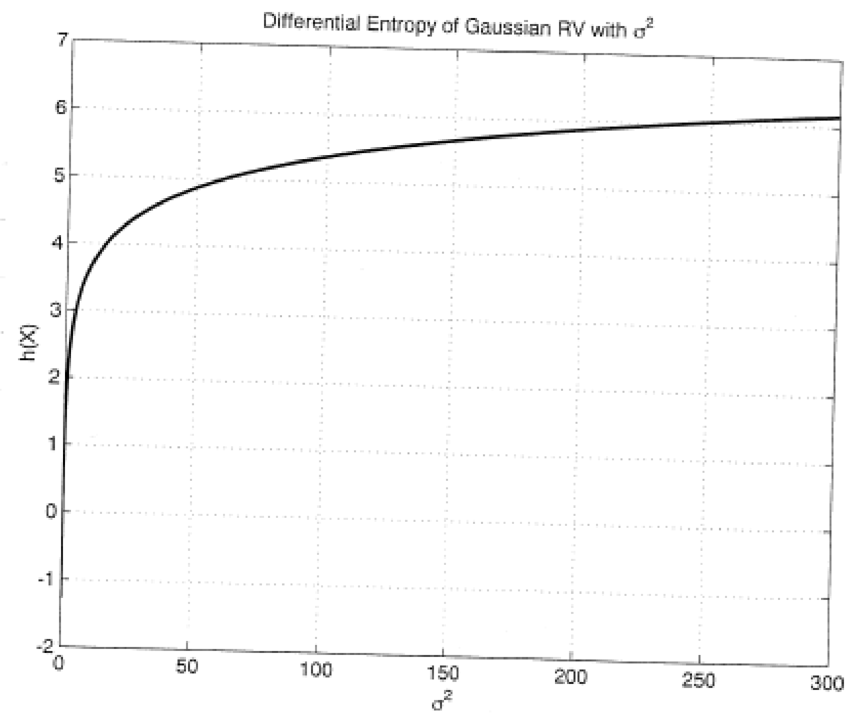
What is the diff Info *h*(*X*) if *X* is a Gaussian RV?

(Let the variance be σ^2 and the mean be μ)

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2} \right]$$

Using this in equation for *h*(*X*) and using properties of log & exp gives:

$$h(X) = \frac{1}{2}\log_2\left[2\pi e\sigma^2\right]$$



Can show that: For <u>any RV</u> *X* w/ variance σ^2

$$h(X) \le \frac{1}{2}\log_2\left[2\pi e\sigma^2\right]$$

That is, the Gaussian has maximum differential entropy!!

Recall Discrete Case: Equally Probable RV gave the max Entropy

Cont. RV results in Info Theory are often derived for the Gaussian case because:

1. Math is easier than for other RVs.
2. Provides an upper bound on $h(X)$ for all other cases.

Avg Mutual Info for Cont. Case

Thus, in general problems of info theory…

$h(X)$ plays the role of $H(X)$ with pretty much the same ideas!!!

But for us it is even more straight-forward… What we are really interested in is $I(X;Y)$ when X is a cont. RV…

So… applying the limiting argument directly:

$$I(X_d;Y_d) = H(X_d) - H(X_d \mid Y_d)$$

We have found this

Need to get this

Can show:

$$H(X \mid Y) = -\sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \Big[f_{X|Y}(x_i \mid y_j) f_Y(y_j) \log_2 \big[f_{X|Y}(x_i \mid y_j) \big] \Big] \Delta\Delta - \log_2 \Delta$$

Note: ***Both*** terms in $I(X_d;Y_d)$ have the pesky $-\log_2\Delta$ term… so they cancel ***BEFORE*** the limit!!!

So… we only have the differential parts left:

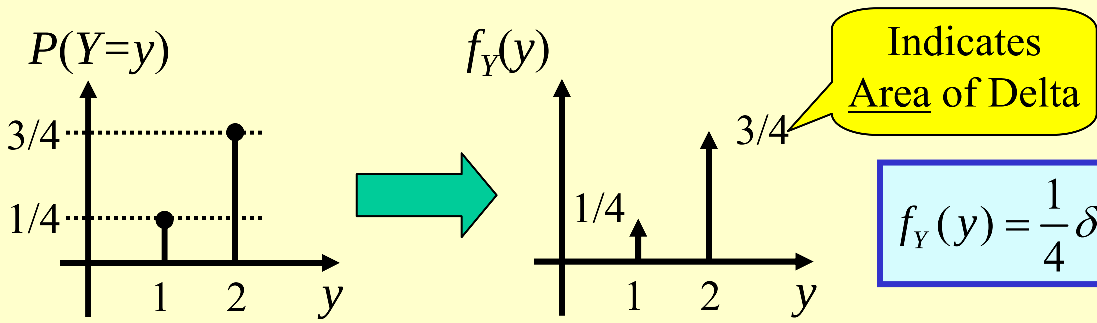
$$I(X;Y) = h(X) - h(X|Y)$$

where
$$h(X|Y) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X|Y}(x|y)\, f_Y(y) \log_2\left[f_{X|Y}(x|y)\right] dxdy$$

Note that this is for X and Y ***both*** continuous RVs… and the result uses PDFs for X and Y

For our case, *Y* is discrete… But this can be handled by using delta functions in a PDF being used to describe a discrete RV

Ex. Let RV *Y* take on discrete values of 1 & 2 with a prob. function $P(Y=1) = 1/4$ and $P(Y=2) = 3/4$



$$f_Y(y) = \frac{1}{4}\delta(y-1) + \frac{3}{4}\delta(y-2)$$

Big Picture Result for Cont. *X* & Disc. *Y*

We use $I(X;Y) = h(X) - h(X|Y)$ as the theoretical minimum rate needed to convey the info that *Y* holds about *X*.

$$I(X;Y) = h(X) - h(X\,|\,Y)$$

- $I(X;Y)$: Amount of Info *Y* conveys about *X*
- $h(X)$: Total Diff. Info that could be conveyed
- $h(X|Y)$: Diff. Info that isn't conveyed

The uncertainty left unresolved about *X*